This document is published in:

Economics Letters, 2014, v. 122, n. 2, pp. 197-202

DOI: 10.1016/j.econlet.2013.11.030

Measurement Error in Imputation Procedures*

Rodolfo G. Campos† Iliana Reggio‡

This version: November 12, 2013

Abstract

We study how estimators that are used to impute consumption in survey data are inconsistent due to measurement error in consumption. Previous research suggests instrumenting consumption to overcome this problem. We show that, if additional regressors are present, then instrumenting consumption may still produce inconsistent estimators due to the likely correlation between additional regressors and measurement error. On the other hand, low correlations between additional regressors and instruments may reduce bias due to measurement error. We apply our findings by revisiting recent research that imputes consumption data from the CEX to the PSID.

JEL classification: C13, C26, E21

Keywords: consumption, measurement error, instrumental variables, Consumer Expenditure Survey, Panel Study of Income Dynamics, income shocks.

*We thank an anonymous referee for very useful comments and are grateful to Richard Blundell and Jesús Carro for their comments on a previous draft of this paper. Magdalena Opazo provided outstanding research assistance. We acknowledge financial support by the Spanish Ministerio de Ciencia y Tecnología (grants ECO2009-13169 and ECO2009-11165) and Ministerio de Economía y Competitividad (grants ECO2012-38134 and ECO2012-31358).

†Department of Economics, IESE Business School, Camino del Cerro del Águila 3, 28023 Madrid, Spain; email: `rcampos@iese.edu`.

‡Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain; email: `ireggio@eco.uc3m.es`.

1 Introduction

In an influential article, Blundell, Pistaferri, and Preston (2008) use a novel procedure to impute consumption data from the Consumer Expenditure Survey (CEX) to the Panel Study of Income Dynamics (PSID). Their procedure consists in estimating a demand for food equation using CEX data and then using the estimated parameters to construct artificial consumption data in the PSID. Although being available only recently, the imputation procedure by Blundell, Pistaferri, and Preston (2008) has been widely adopted.[1]

A difference with previous imputation procedures, most notably that of Skinner (1987), is the employment of instrumental variable techniques to deal with measurement error.[2] Blundell, Pistaferri, and Preston (2004, 2008) argue that, despite measurement error, the sensitivity of food consumption to total consumption can be consistently estimated if total consumption is instrumented. They also show that consistency leads to the sample variance of imputed consumption to converge in probability to the same limit as the variance of true consumption, up to an additive term, implying that the variance of imputed consumption is just an upward translated version of the variance of true consumption with the same time trends. Therefore, consumption inequality measured as the variance of imputed consumption is a useful indicator for changes in true consumption inequality over time.

Blundell, Pistaferri, and Preston (2004) consider a special case in which total consumption is the only variable on the right-hand side of the demand for food equation and prove that, if total consumption is instrumented, then the coefficient on this variable is consistently estimated even in the presence of measurement error. In practice, however, a demand for food will have additional covariates, such as prices and demographic characteristics.

Are the results of Blundell, Pistaferri, and Preston (2004, 2008) robust to the inclusion of additional covariates? In this note we argue that the answer is 'no'. In Section 2 we

[1] Examples include the work by Guvenen and Smith (2010), Hryshko, Luengo-Prado, and Sorensen (2010), Attanasio, Hurst, and Pistaferri (2012), and Michelacci and Ruffo (2013). Others, such as Kaplan and Violante (2010), Abraham, Koehne, and Pavoni (2012), and Broer (2012) have implicitly adopted the imputation procedure by directly using the original imputed data from Blundell, Pistaferri, and Preston (2008), which is available online at `http://www.aeaweb.org/aer/data/dec08/20050545_data.zip`.

[2] Measurement error is a pervasive problem in consumption surveys, particularly in those using recall methods. See, for example, Ahmed, Brzozowski, and Crossley (2006) and Battistin and Padula (2010).

consider the case in which the demand for food includes another variable in addition to total consumption and show that measurement error in consumption biases the estimates through its correlation with the additional covariates. Moreover, because the coefficient on total consumption fails to be consistently estimated, the variance of imputed consumption does not replicate the movement of the variance of true consumption. On the other hand, inspection of the expression for the asymptotic bias due to measurement error reveals that, if the additional covariates are orthogonal to the instrument, then the coefficient on total consumption can be consistently estimated, and the variance of imputed consumption may recover its useful property. Finally, in Section 3 we provide a practical application of our results by revisiting the demand for food estimation of Blundell, Pistaferri, and Preston (2008).

2 IV estimation in presence of an additional covariate

If the demand equation for food of Blundell, Pistaferri, and Preston (2004, 2008) is augmented by an additional variable, it takes the form

$$f_i = \beta_0 + \gamma c_i + \beta_1 d_i + e_i. \tag{1}$$

Demand for food is a function that relates expenditure on food f (usually in logs) to total non-durable expenditure c (also usually in logs) and another variable d representing, for example, the price of food, the price of substitutes or complementary goods, or a characteristic of the household that acts as a demand shifter. The parameter γ measures the sensitivity of food consumption to total consumption (the budget elasticity, if variables are measured in logs). Unobserved heterogeneity is represented by e. The single departure from the specification of Blundell, Pistaferri, and Preston (2004) is that the variable d is included.

Following Blundell, Pistaferri, and Preston (2004), measured consumption expenditure c_i^* equals the sum of true consumption expenditure and a measurement error term: $c_i^* = c_i + u_i$. The demand equation expressed in terms of c_i^* is

$$f_i = \beta_0 + \gamma c_i^* + \beta_1 d_i + e_i - \gamma u_i. \tag{2}$$

Consistent estimation of the parameters in this equation is hindered by a potentially non-zero covariance of c^* and d with measurement error u.

Imputation proceeds by using the parameters from this equation estimated with CEX data together with observations of f and d from the PSID to obtain predicted consumption observations for all the households in the PSID. After inverting the demand for food, imputed consumption is calculated as

$$\hat{c}_i = \frac{1}{\hat{\gamma}} \left[f_i - \hat{\beta}_0 - \hat{\beta}_1 d_i \right]. \tag{3}$$

It is well known that in the presence of classical errors-in-variables (CEV), OLS estimators from the food-demand equation (2) are inconsistent.[3] Blundell, Pistaferri, and Preston (2004) prove that instrumenting c^* with a valid instrument z eliminates any asymptotic bias and yields consistent estimators if total consumption c^* is the sole regressor. In Proposition 1 we derive how the presence of the additional covariate d affects the probability limits of $\hat{\gamma}$ and $\hat{\beta}_1$, and the relationship between the variance of imputed consumption $V(\hat{c})$ and the variance of true consumption $V(c)$.[4]

Proposition 1

Let z be a valid instrument for c^*, d an exogenous regressor in (1), $\hat{\beta}_1$ the IV estimator of β_1, and $\hat{\gamma}$ the IV estimator of γ. Then, the IV estimation of (2) yields the following asymptotic results

$$\text{plim}\, \hat{\gamma} = \gamma \left[1 + \frac{Cov(d,u)Cov(d,z)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)} \right] \tag{4}$$

$$\text{plim}\, \hat{\beta}_1 = \beta_1 - \gamma \frac{Cov(c^*,z)Cov(d,u)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)} \tag{5}$$

[3] See, for example, Wooldridge (2002, Ch. 4).

[4] The proof is in the Appendix.

$$
\begin{aligned}
\operatorname{plim} V(\hat{c}) = \left(\frac{1}{1 + \frac{Cov(d,u)Cov(d,z)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)}} \right)^2 \Bigg[& \operatorname{plim} V(c) \\
& + \frac{1}{\gamma^2} \operatorname{plim} V(e) + \frac{2}{\gamma} \operatorname{plim} Cov(e,c) \\
& + \left(\frac{Cov(c^*,z)Cov(d,u)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)} \right)^2 \operatorname{plim} V(d) \\
& + 2 \left(\frac{Cov(c^*,z)Cov(d,u)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)} \right) \operatorname{plim} Cov(c,d) \Bigg] \qquad (6)
\end{aligned}
$$

In contrast to what happens in the absence of an additional regressor d, both estimators are inconsistent despite z being a valid instrument for c^*. To ensure consistent estimators, the additional variable d (i.e., any additional variable that belongs in the demand for food) would need to be instrumented as well. Because it is not, asymptotic bias due to measurement error sneaks back into the estimates through the covariance between the additional variable and measurement error, $Cov(d, u)$. This covariance is potentially non-zero, as implied by studies linking measurement error to demographic characteristics that usually appear in the demand for food.

Whereas the classical study of measurement error in survey data by Bound, Brown, and Mathiowetz (2001) does not discuss measurement error in the context of consumption expenditure data, recent research on consumer surveys has uncovered evidence suggesting a correlation between measurement error in consumption and demographic characteristics. Because it is usually a single respondent who is asked to recall expenditures for the whole household, measurement error is found to be correlated with household size (Gibson, 2002). Moreover, when consumption is measured in logs, Ahmed, Brzozowski, and Crossley (2006, Table 8) find statistically significant levels of correlation between measurement error and other household composition variables, such as the number of children, youths, and seniors. Furthermore, a host of studies have argued that measurement error is correlated with income (e.g., Bee, Meyer, and Sullivan, 2013; Sabelhaus, Johnson, Ash, Swanson, Garner, Greenlees, and Henderson, 2013); if income is not among the covariates, then part of this correlation will be picked up by demographic variables correlated with income, such as age and education. In fact, using Swedish car registration data that can be matched to individual survey respondents, Koijen, Nieuwerburgh, and Vestman (2013) show that measurement error is correlated with education even when age, income, and wealth are used as controls.

We are not aware of any study focusing directly on the correlation between measurement error in consumption and prices, the other set of variables in a demand equation. However, it has been found that consumption expenditures in the CEX diverge over time from Personal Consumption Expenditure (PCE). For example, Passero, Garner, and McCully (2013) measure that the ratio of CEX to PCE consumption has fallen by ten percentage points over a 18-year period. Because the gap between CEX and PCE consumption is a measure of measurement error, its time trend implies statistical correlation between measurement error and prices, which do also have a time trend. Moreover, the cyclical discrepancies between the CEX and PCE documented by Campos, Reggio, and García-Píriz (2013) could be indicative of measurement error that is cyclical. Given that prices tend to be cyclical as well, this provides an additional channel through which measurement error may be correlated with prices.

In comparison to the result by Blundell, Pistaferri, and Preston (2004), the presence of the additional regressor d augments the expression for the variance of imputed consumption in (6) by two additive terms. Because of measurement error, the variance of imputed consumption is additively impacted by the variance of the additional regressor $V(d)$ and the covariance $Cov(c, d)$. More importantly, the multiplicative term in front of $\operatorname{plim} V(c)$ is different from one because $\hat{\gamma}$ is inconsistent even if consumption is instrumented. Thus, the variance of predicted consumption does not move in lockstep with the variance of true consumption, implying that the evolution of the variance of true consumption over time is not tracked by the variance of imputed consumption.

Close inspection of the expressions in the proposition reveals that there is a way to salvage the result that the sensitivity of food consumption to total consumption is consistently estimated. Notice that if $Cov(d, z) = 0$, implying that the instrument for consumption expenditure is orthogonal to the additional regressor, then the estimator $\hat{\gamma}$ is consistent. In turn, the consistency of $\hat{\gamma}$ implies that the slope coefficient in the expression for the variance in the proposition is one. Furthermore, if the assumption of Blundell, Pistaferri, and Preston (2004) that in (6) the sum of the first two additive terms is time stationary is extended to the sum of all four terms, then this implies that the variance of true consumption is again tracked by the variance of imputed consumption.[5]

[5]Whether time stationarity can be reasonably expected to hold will, of course, depend on the par-

Whether $\hat{\gamma}$, the estimator of the sensitivity of food consumption to total consumption, is consistent depends on whether the orthogonality condition $Cov(d, z) = 0$ is fulfilled. Strict fulfillment will be impossible in practice, therefore leading to some amount of bias. In unreported Monte Carlo simulations we found that non-zero correlations between the instrument z and the covariate d may produce considerable bias; correlations of around 0.3 may be enough to drive an important wedge between γ and $\hat{\gamma}$ (Campos and Reggio, 2012). In comparison, the correlation between several regressors and the instrument in the data of Blundell, Pistaferri, and Preston (2008) surpasses 0.4 and, in one case, exceeds 0.6 (the correlation with being a high school graduate).

3 Application

To implement their imputation procedure, Blundell, Pistaferri, and Preston (2008) estimate a demand equation for food using CEX data from 1980 to 1992. They instrument total nondurable consumption with the average of the hourly wage of the husband (by cohort, year, and education) and the average of the hourly wage of the wife (also by cohort, year, and education).

In light of our results, a potential problem with the estimation arises if instruments are correlated with the additional regressors they use. In their specification there are two groups of variables with high sample correlations with the instruments: education dummies and prices. In the first column of Table 1 we report the correlations of these variables with the average of the hourly wage of the husband. The correlation between the hourly wage of the husband and education dummies are around 0.6; the correlations with prices are somewhat lower.

There are two complementary ways of reducing measurement error bias. One way is to drop the problematic regressors in the estimation of the budget elasticity, and the other is to use an alternative instrument that is less correlated with the regressors. We show examples of both approaches.

As a benchmark, we replicate the specification of Blundell, Pistaferri, and Preston (2008). The budget elasticity is estimated to be 0.850 (Table 2, Col. 1). If education

ticular context. However, even in contexts in which the assumption of time stationarity is questionable, a solution (at the cost of possibly introducing omitted variable bias) is to drop suspect covariates from the regression altogether. We thank an anonymous referee for pointing this out.

Table 1: *Correlations between the hourly wage of the husband and selected regressors.*

	BPP	Alt. IV	BPP - Real	Alt. IV - Real
Prices				
- Food	0.467	0.753	-0.072	-0.150
- Alcohol and Tobacco	0.469	0.758	-0.069	-0.134
- Fuel and Utilities	0.427	0.696	-0.066	-0.117
- Transports	0.455	0.738	-0.078	-0.148
Education				
- Elementary	-0.567	-0.058	-0.607	-0.013
- HS Graduate	0.627	0.055	0.674	0.010

Correlations of regressors with the largest correlations with the average hourly wage of the husband. The label "BPP" indicates the use of the instruments of Blundell, Pistaferri, and Preston (2008) whereas "Alt. IV" indicates an alternative definition of the instruments. "Real" indicates the use of deflated data.

dummies are excluded from the baseline specification, then a lower budget elasticity of 0.799 is obtained (Table 2, Col. 3). It can be argued that this is not a fair comparison because the exclusion of education dummies also implies dropping the interactions between education and $\ln c$. Thus, we also re-estimate the specification of Blundell, Pistaferri, and Preston (2008) without the interactions. The result is stronger. Comparing columns 2 and 3 in Table 2, we find that the estimated budget elasticity drops from 1.081 to 0.799 if education dummies are excluded.

The difference in the results when education is removed suggests that measurement error could be biasing the estimate of the budget elasticity upward. It is far from a definitive proof; if education dummies are deemed necessary in the demand equation, then their removal may generate omitted variable bias. On the other hand, the demand function contains total consumption expenditure that is instrumented by wage rates. This reduces the role of education as a proxy for income. In any case, the sensitivity of the estimate of the budget elasticity to the removal of education dummies should at least cast doubt on the exact value of the estimate.

The second approach does not require dropping any variables from the demand equation. The difference is in the construction of the instruments. To achieve less correlation with education we calculate the average hourly wage of the husband and the average hourly wage of the wife by cohort and year but without conditioning on education. Doing so lowers the correlation between the instrument and education dummies to close to zero

(Table 1, Col. 2). Using these alternative instruments, the point estimate of the budget elasticity drops to 0.718 (Table 2, Col. 4). In this case, the relevant comparison is with the original estimate of 0.850. The estimate is less precise and does not allow to statistically distinguish between these values at the usual probability thresholds. Nevertheless, if the difference in the point estimates is attributed to measurement error, then the evidence indicates that the budget elasticity is biased upward, as before.

Table 2: *Sensitivity of the budget elasticity to different specifications and to the use of alternative instruments.*

VARIABLES	(1) BPP	(2) No Interactions	(3) No Education	(4) Alt. IV
$\ln c$	0.850***	1.081***	0.799***	0.718***
	(0.151)	(0.112)	(0.032)	(0.203)
$\ln c$ x HS	0.073			-0.004
	(0.072)			(0.076)
$\ln c$ x College	0.083			0.058
	(0.089)			(0.108)
Observations	14,430	14,430	14,430	14,430
R-squared	0.671	0.619	0.687	0.682
RMSE	0.249	0.268	0.243	0.245

*Standard errors in parentheses (*** p < 0.01). In the first three columns the instruments are the average (by cohort, year, and education) of the hourly wage of the husband and the average (also by cohort, year, and education) of the hourly wage of the wife. In column (4) the instruments are the average (by cohort and year) of the hourly wage of the husband and the average (also by cohort and year) of the hourly wage of the wife.*

The other group of variables correlated with the instruments are prices. Their correlation with the instrument is not removed by the alternative definition of the instrument; in fact, correlations with prices are higher (Table 1, Col. 2). The reason behind the large correlation with prices is that total consumption expenditure enters the food demand equation in nominal terms. Wages used to instrument consumption are also nominal. Wages and prices are linked by inflation.

A more flexible specification for the demand for food breaks this link. We separate nominal expenditures into a real component and a price index, and do the same with the instrument. We do so by deflating nominal values using the Consumer Price Index and add this index as an additional regressor. This change in the specification reduces the correlation between the instrument and additional regressors. Correlations with prices are lower both for the original instrument in Blundell, Pistaferri, and Preston (2008) (Table 1, Col. 3) and for the alternative definition of the instrument (Table 1, Col. 4).

We repeat our previous regressions using real consumption expenditure instrumented by real wages. Results are shown in Table 3. The columns are analogous to those in Table 2. The specification of Blundell, Pistaferri, and Preston (2008) with real expenditures and real wages produces an estimate of 0.937 (Table 2, Col. 1). Again, lower point estimates for the budget elasticity are obtained when education dummies are dropped (Table 2, Col. 3) and when an alternative definition is used for the instruments (Table 2, Col.4).

Table 3: *Sensitivity of the budget elasticity to different specifications and the use of alternative instruments using real expenditures instrumented by real wages.*

VARIABLES	(1) BPP	(2) No Interactions	(3) No Education	(4) Alt.IV
$\ln c$	0.937***	1.025***	0.786***	0.772***
	(0.119)	(0.100)	(0.032)	(0.211)
$\ln c$ x HS	0.112			-0.101
	(0.129)			(0.129)
$\ln c$ x College	0.018			-0.151
	(0.121)			(0.126)
Observations	14,430	14,430	14,430	14,430
R-squared	0.655	0.635	0.686	0.682
RMSE	0.255	0.262	0.243	0.245

*Standard errors in parentheses (*** $p < 0.01$). All regressions use consumption and wages in real terms. In the first three columns the instruments are the average (by cohort, year, and education) of the hourly wage of the husband and the average (also by cohort, year, and education) of the hourly wage of the wife. In column (4) the instruments are the average (by cohort and year) of the hourly wage of the husband and the average (also by cohort and year) of the hourly wage of the wife.*

In conclusion, our results in this section indicate that in the estimation of the demand for food, the instruments used by Blundell, Pistaferri, and Preston (2008) are highly correlated with two sets of additional regressors, education and prices. Modifying the specification and using alternative instruments are two approaches that lead to lower correlations between those instruments and the additional regressors. Estimated coefficients of the budget elasticity were lower in all cases, suggesting that the unobserved correlation between additional regressors and measurement error is leading to an overestimation of this elasticity.

Blundell, Pistaferri, and Preston (2008) use the imputation procedure only as an intermediate step. The final objective is to estimate the response of household consumption to permanent and transitory income shocks. In Table 4 we show how their answers are influenced by the different demands for food we estimated.[6]

The response of consumption to transitory income shocks (denoted by ψ) is not substantially affected by the different imputation procedures. It remains low and is not significantly different from zero. In contrast, either dropping education dummies or using the alternative definition of the instruments, leads to a rise in ϕ, the response of consumption to permanent shocks. This happens regardless of whether nominal or real expenditures are used for the imputation. This suggests that the likely bias in the budget elasticity implies an underestimation of the impact of permanent income shocks on consumption.[7]

[6] For details on the model used to estimate the response to permanent and transitory income shocks consult Blundell, Pistaferri, and Preston (2008). We obtain our results by using their data and adapting their code.

[7] On the other hand, the values estimated for ϕ are inside the range of values considered by Blundell, Pistaferri, and Preston (2008) in their robustness checks.

Table 4: *Robustness of the response to permanent and transitory income shocks.*

	BPP	No education	Alt. IV
Nominal Imputation			
ϕ	0.6423	0.7882	0.8186
	(0.0945)	(0.1153)	(0.1191)
ψ	0.0533	0.0558	0.0601
	(0.0435)	(0.0523)	(0.0584)
Real Imputation			
ϕ	0.5988	0.7871	0.7668
	(0.0877)	(0.1150)	(0.1106)
ψ	0.0453	0.0545	0.0501
	(0.0396)	(0.0519)	(0.0553)

Estimation results from using the alternative imputation strategies discussed in the main text. Standard errors in parentheses.

4 Conclusion

In this note we have shown that the presence of measurement error may produce inconsistent estimators in procedures used to impute consumption even if instrumental variables are used. By explicitly deriving the expression for the asymptotic bias due to measurement error we were able to detect that if the instrument is orthogonal to the additional regressors, then problems due to measurement error are mitigated. Since the orthogonality condition refers to observable variables, practitioners can check whether it is satisfied whenever they use the imputation procedure by Blundell, Pistaferri, and Preston (2008).

Our application revisiting the work by Blundell, Pistaferri, and Preston (2008) provides an example of the type of robustness checks that can be performed on the imputation procedure. Our results tended to yield lower estimates for the budget elasticity of food consumption. In turn, in the context of the model by Blundell, Pistaferri, and Preston (2008), these revised estimates imply a larger role of permanent income shocks in driving consumption.

References

Abraham, A., S. Koehne, and N. Pavoni (2012): "Optimal income taxation with asset accumulation," MPRA Paper 38629, University Library of Munich, Germany.

Ahmed, N., M. Brzozowski, and T. F. Crossley (2006): "Measurement Errors in Recall Food Consumption Data," *The Institute for Fiscal Studies*, 06/21.

Attanasio, O., E. Hurst, and L. Pistaferri (2012): "The Evolution of Income, Consumption, and Leisure Inequality in The US, 1980-2010," NBER Working Papers 17982, National Bureau of Economic Research, Inc.

Battistin, E., and M. Padula (2010): "Survey Instruments and the Reports of Consumption Expenditures: Evidence from the Consumer Expenditure Surveys," CSEF Working Papers 259, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.

Bee, A., B. D. Meyer, and J. X. Sullivan (2013): "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?," in *Improving the Measurement of Consumer Expenditures*, ed. by C. Carroll, T. Crossley, and J. Sabelhaus. University of Chicago Press.

Blundell, R., L. Pistaferri, and I. Preston (2004): "Imputing Consumption in the PSID using food demand estimates from the CEX.," *Institute for Fiscal Studies Working Paper*, 04/27.

——— (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement error in survey data," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5 of *Handbook of Econometrics*, chap. 59, pp. 3705–3843. Elsevier.

Broer, T. (2012): "The wrong shape of insurance? What cross-sectional distributions tell us about models of consumption-smoothing," mimeo.

Campos, R. G., and I. Reggio (2012): "Measurement error and imputation of consumption in survey data," Economics Working Papers we1219, *Universidad Carlos III*.

Campos, R. G., I. Reggio, and D. García-Píriz (2013): "Micro versus macro consumption data: the cyclical properties of the Consumer Expenditure Survey," *Applied Economics*, 45(26), 3778–3785.

Gibson, J. (2002): "Why Does the Engel Method Work? Food Demand, Economies of Size and Household Survey Methods," *Oxford Bulletin of Economics and Statistics*, 64(4), 341–59.

Guvenen, F., and A. Smith (2010): "Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach," NBER Working Papers 16327, National Bureau of Economic Research, Inc.

Hryshko, D., M. J. Luengo-Prado, and B. E. Sorensen (2010): "House prices and risk sharing," *Journal of Monetary Economics*, 57(8), 975–987.

Kaplan, G., and G. L. Violante (2010): "How Much Consumption Insurance beyond Self-Insurance?," *American Economic Journal: Macroeconomics*, 2(4), 53–87.

Koijen, R., S. V. Nieuwerburgh, and R. Vestman (2013): "Judging the Quality of Survey Data by Comparison with "Truth" as Measured by Administrative Records: Evidence From Sweden," in *Improving the Measurement of Consumer Expenditures*, ed. by C. Carroll, T. Crossley, and J. Sabelhaus. University of Chicago Press.

Michelacci, C., and H. Ruffo (2013): "Optimal Life Cycle Unemployment Insurance," mimeo, CEMFI.

Passero, W., T. I. Garner, and C. McCully (2013): "Understanding the Relationship: CE Survey and PCE," in *Improving the Measurement of Consumer Expenditures*, ed. by C. Carroll, T. Crossley, and J. Sabelhaus. University of Chicago Press.

Sabelhaus, J., D. Johnson, S. Ash, D. Swanson, T. Garner, J. Greenlees, and S. Henderson (2013): "Is the Consumer Expenditure Survey Representative by Income?," in *Improving the Measurement of Consumer Expenditures*, ed. by C. Carroll, T. Crossley, and J. Sabelhaus. University of Chicago Press.

Skinner, J. (1987): "A superior measure of consumption from the Panel Study of Income Dynamics," *Economics Letters*, 23, 213–216.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

A Proof of Proposition 1

Estimators

Apply the standard IV formula $\hat{\theta} = (Z^\top X)^{-1} Z^\top y$, and properties of convergence in probabilities, to obtain the probability limit of both estimators: $\hat{\beta}_1$ and $\hat{\gamma}$. The plim of the estimator of the parameter of the variable measured with error is

$$\text{plim}\,\hat{\gamma} = \frac{Cov(f,z)V(d) - Cov(d,z)Cov(d,y)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)} \tag{A.7}$$

Replacing equation (2) in (A.7) yields

$$\begin{aligned} \text{plim}\,\hat{\gamma} = \frac{1}{\Phi} \{ & V(d) \left[\beta_1 Cov(d,z) + \gamma Cov(c^*,z) + Cov(e,z) - \gamma Cov(u,z)\right] \\ & - Cov(d,z) \left[\beta_1 V(d) + \gamma Cov(d,c^*) + Cov(d,e) - \gamma Cov(d,u)\right] \} \end{aligned} \tag{A.8}$$

where $\Phi \equiv V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)$. After some algebra the probability limit of $\hat{\gamma}$ is

$$\begin{aligned} \text{plim}\,\hat{\gamma} = \gamma & + \frac{Cov(e,z)V(d)}{\Phi} - \frac{\gamma Cov(u,z)V(d)}{\Phi} \\ & - \frac{Cov(d,z)Cov(d,e)}{\Phi} + \frac{\gamma Cov(d,z)Cov(d,u)}{\Phi} \end{aligned} \tag{A.9}$$

Because z is a valid instrument, $Cov(z,e) = Cov(z,u) = 0$, and because d is exogenous in (1), $Cov(d,e) = 0$. Therefore,

$$\text{plim}\,\hat{\gamma} = \gamma \left[1 + \frac{Cov(u,d)Cov(d,z)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)}\right]. \tag{A.10}$$

A similar derivation is done for β_1; the IV formula for $\hat{\beta}_1$ implies that the probability limit is

$$\text{plim}\,\hat{\beta}_1 = \frac{Cov(d,f)Cov(c^*,z) - Cov(c^*,x_1)Cov(f,z)}{V(d)Cov(c^*,z) - Cov(c^*,d)Cov(d,z)} \tag{A.11}$$

Replace f using equation (2) to obtain

$$\begin{aligned} \text{plim}\,\hat{\beta}_1 = \frac{1}{\Phi} \{ & Cov(c^*,z) \left[\beta_1 V(d) + \gamma Cov(c^*,d) + Cov(d,e) - \gamma Cov(d,u)\right] \\ & - Cov(c^*,d) \left[\beta_1 Cov(d,z) + \gamma Cov(c^*,z) + Cov(e,z) - \gamma Cov(u,z)\right] \} \end{aligned} \tag{A.12}$$

After some algebra, the probability limit of $\hat{\beta}_1$ is

$$\operatorname{plim}\hat{\beta}_1 = \beta_1 - \frac{Cov(c^*, d)\left[Cov(e, z) - \gamma Cov(u, z)\right]}{V(d)Cov(c^*, z) - Cov(c^*, d)Cov(d, z)} + \frac{Cov(c^*, z)\left[Cov(d, e) - \gamma Cov(d, u)\right]}{V(d)Cov(c^*, z) - Cov(c^*, d)Cov(d, z)} \tag{A.13}$$

If z is a valid instrument, then $Cov(e, z) = Cov(u, z) = 0$. If d is exogenous in (1), then $Cov(d, e) = 0$. Therefore,

$$\operatorname{plim}\hat{\beta}_1 = \beta_1 - \gamma\frac{Cov(c^*, z)Cov(d, u)}{V(d)Cov(c^*, z) - Cov(c^*, d)Cov(d, z)} \tag{A.14}$$

Variance of imputed consumption

Use (1) to substitute for f_i in (3) to obtain the relationship between $\hat{c}_i$ and c_i:

$$\hat{c}_i = \frac{1}{\hat{\gamma}}\left[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)d_i + \gamma c_i + e_i\right]. \tag{A.15}$$

Applying properties of convergence in probability to (A.15), write the probability limit of the sample variance of predicted consumption as

$$\begin{aligned}\operatorname{plim}V(\hat{c}_i) = {} & \left(\frac{\gamma}{\operatorname{plim}\hat{\gamma}}\right)^2 \operatorname{plim}V(c) + \left(\frac{\beta_1 - \operatorname{plim}\hat{\beta}_1}{\operatorname{plim}\hat{\gamma}}\right)^2 \operatorname{plim}V(d) \\ & + \left(\frac{1}{\operatorname{plim}\hat{\gamma}}\right)^2 \operatorname{plim}V(e) + 2\left(\frac{\gamma}{(\operatorname{plim}\hat{\gamma})^2}\right)\operatorname{plim}Cov(e, c) \\ & + 2\gamma\left(\frac{\beta_1 - \operatorname{plim}\hat{\beta}_1}{(\operatorname{plim}\hat{\gamma})^2}\right)\operatorname{plim}Cov(c, d) \\ & + 2\left(\frac{\beta_1 - \operatorname{plim}\hat{\beta}_1}{(\operatorname{plim}\hat{\gamma})^2}\right)\operatorname{plim}Cov(e, d)\end{aligned} \tag{A.16}$$

Use (A.10) and (A.14) to substitute $\operatorname{plim}\hat{\gamma}$ and $\beta_1 - \operatorname{plim}\hat{\beta}_1$ in (A.16) and obtain the expression for the probability limit of the sample variance of predicted consumption:

$$\begin{aligned}
\operatorname{plim} V(\hat{c}_i) = &\left(\frac{1}{1+\frac{Cov(u,d)Cov(d,z)}{V(d)Cov(c^*,z)-Cov(c^*,d)Cov(d,z)}}\right)^2 \Bigg[\operatorname{plim} V(c) + \frac{1}{\gamma^2}\operatorname{plim} V(e) \\
&+ \left(\frac{Cov(c^*,z)Cov(d,u)}{V(d)Cov(c^*,z)-Cov(c^*,d)Cov(d,z)}\right)^2 \operatorname{plim} V(d) \\
&+ 2\left(\frac{Cov(c^*,z)Cov(d,u)}{V(d)Cov(c^*,z)-Cov(c^*,d)Cov(d,z)}\right) \operatorname{plim} Cov(c,d) \\
&+ \frac{2}{\gamma}\operatorname{plim} Cov(e,c)\Bigg]
\end{aligned} \tag{A.17}$$